

Human T-Cell Leukemia Virus Type 1 Integration Target Sites in the Human Genome: Comparison with Those of Other Retroviruses^{∇‡}

David Derse,^{1†} Bruce Crise,^{2†} Yuan Li,³ Gerald Princler,¹ Nicole Lum,⁴ Claudia Stewart,⁴
Connor F. McGrath,⁵ Stephen H. Hughes,¹ David J. Munroe,⁴ and Xiaolin Wu^{4*}

*HIV Drug Resistance Program, NCI-Frederick, Frederick, Maryland 21702,¹ and Gene Expression Laboratory,²
AIDS Vaccine Program,³ Laboratory of Molecular Technology,⁴ and Target Structure-Based Drug Discovery Group,⁵
SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland 21702*

Received 14 December 2006/Accepted 29 March 2007

Retroviral integration into the host genome is not entirely random, and integration site preferences vary among different retroviruses. Human immunodeficiency virus (HIV) prefers to integrate within active genes, whereas murine leukemia virus (MLV) prefers to integrate near transcription start sites and CpG islands. On the other hand, integration of avian sarcoma-leukosis virus (ASLV) shows little preference either for genes, transcription start sites, or CpG islands. While host cellular factors play important roles in target site selection, the viral integrase is probably the major viral determinant. It is reasonable to hypothesize that retroviruses with similar integrases have similar preferences for target site selection. Although integration profiles are well defined for members of the lentivirus, spumaretrovirus, alpharetrovirus, and gammaretrovirus genera, no members of the deltaretroviruses, for example, human T-cell leukemia virus type 1 (HTLV-1), have been evaluated. We have mapped 541 HTLV-1 integration sites in human HeLa cells and show that HTLV-1, like ASLV, does not specifically target transcription units and transcription start sites. Comparing the integration sites of HTLV-1 with those of ASLV, HIV, simian immunodeficiency virus, MLV, and foamy virus, we show that global and local integration site preferences correlate with the sequence/structure of virus-encoded integrases, supporting the idea that integrase is the major determinant of retroviral integration site selection. Our results suggest that the global integration profiles of other retroviruses could be predicted from phylogenetic comparisons of the integrase proteins. Our results show that retroviruses that engender different insertional mutagenesis risks can have similar integration profiles.

Integration of the viral DNA genome into the host cell genome is a necessary step for retrovirus replication (14). Retroviral integration site selection not only is central to the biology of retroviruses but also is important for gene therapy, because retroviral vectors are widely used for gene delivery and the risk of insertional mutagenesis is real (9, 22, 23). The integration process is catalyzed by the viral integrase and involves the cleavage and joining of viral and host DNA (14). Early studies showed that most of the host genome is accessible for retroviral integration but that target site selection is not totally random (14, 28, 46, 47, 55). Availability of the sequence of the human genome has enabled large-scale studies of retroviral integration sites (2, 13, 17, 25, 34, 40–42, 49, 52, 57). The most surprising finding is that retroviruses from diverse genera have different target site preferences. For example, human immunodeficiency virus (HIV) has a strong preference for integration into genes or transcription units (49, 57). In contrast, murine leukemia virus (MLV) prefers to integrate near transcription start sites or CpG island regions (57). Avian sarcoma-leukosis virus (ASLV), on the other hand, has a much

weaker preference for any of these specific locations (40, 41). Both a simian immunodeficiency virus (SIV)/SIV-based vector and a feline immunodeficiency virus (FIV)-based vector were reported to have patterns of integration that are very similar to those for HIV (17, 25, 27). The integration sites for a foamy virus (FV)-based vector in human cells (42, 52) showed that the virus has a preference for integrating near transcription start sites or CpG islands (52), which is similar to the case for MLV. Although the target site selection is not strongly sequence specific, weak palindromic consensus sequences have been identified at the integration sites of many retroviruses (26, 58).

Cellular cofactors may play important roles in retroviral integration site selection (8, 56). Lens epithelium-derived growth factor (LEDGF/p75) has been shown to bind to HIV integrase (11, 35–37, 53) and to contribute to HIV's preference for integrating into genes (12). Lewinski et al. recently showed that integrase is the principal viral determinant in target site selection (34). In that study, a chimeric HIV virus with an MLV integrase integrated with a target site specificity similar to that of MLV. This suggests that retroviruses with similar integrases should have similar target site preference.

Human T-cell leukemia virus type 1 (HTLV-1), a member of the deltaretrovirus genus, is the causative agent of adult T-cell leukemia and HTLV-1 associated myelopathy/tropical spastic paraparesis (39, 54). HTLV-1 differs from the retroviruses described above, and it provides an opportunity to test the relationship between integrase phylogeny and integration site selection. Although the viral Tax protein is clearly involved in

* Corresponding author. Mailing address: Laboratory of Molecular Technology, SAIC-Frederick, Inc., NCI-Frederick, 915 Toll House Avenue, Frederick, MD 21702. Phone: (301) 846-7677. Fax: (301) 846-6100. E-mail: forestwu@mail.nih.gov.

† These authors contributed equally to this work.

‡ Supplemental material for this article may be found at <http://jvi.asm.org/>.

[∇] Published ahead of print on 4 April 2007.

oncogenic transformation, it is still unclear whether HTLV-1 integration sites influence the expression of cellular or viral genes that relate to the development of disease. There have been several studies of HTLV-1 integration sites in the human genome (19, 24, 31, 32, 44), but the number of sites examined in the majority of these studies was small and the integration sites were cloned from chronically infected patients. In one study, Doi et al. characterized 56 HTLV-1 integration sites from carrier cells and 59 sites from leukemia cells (19) and found that in carrier cells, HTLV-1 integration tended to occur in heterochromatin aliphoid repeated regions, whereas in leukemia cells, HTLV-1 integration favored actively transcribed genes. This difference may arise from the different selection pressures in carrier versus leukemic cells after virus integration.

Here we examine HTLV-1 integration sites in HeLa cells infected with HTLV-1 vectors that express a reporter gene but no viral proteins. A total of 541 HTLV-1 integration sites were cloned and sequenced from acutely infected HeLa cells, analyzed, and compared with the integration sites for five other retroviruses (ASLV, FV, MLV, SIV, and HIV) in relation to currently available genomic features. Our results show that HTLV-1 integrates into the human genome with little preference for most of the genomic features analyzed, which is similar to the case with ASLV. The integration preferences for the six retroviruses can be separated into three distinct groups based on cluster analysis of integration site preferences. In both the cluster analysis of integration site preference and phylogenetic analysis of integrase proteins, SIV was most similar to HIV and formed one group, FV was most similar to MLV and formed a second group, and HTLV-1 was most similar to ASLV, forming a third group.

MATERIALS AND METHODS

Cloning of HTLV-1 integration sites in HeLa cells. HTLV-1 virus-like particles were prepared by transfecting 293T cells with the packaging plasmid pCMV-HT1 plus the transfer vector pHTC-neo. The HTLV-1 vectors are identical to those described previously except that the transfer vector used here contains the neomycin resistance gene instead of luciferase (18). HeLa cells were used for the infection, because previous studies from our group and other groups have used these cells to investigate the integration profiles of MLV, HIV, and ASLV (34, 41, 57). After overnight transfection, 293T cells were washed and then treated with mitomycin C (10 μ g/ml) for 4 h to prevent further cell division. The 293T cells were washed again and mixed with target HeLa cells for cocultivation. Three days later, the cells were diluted, replated, and grown in the presence of G418 (200 μ g/ml) to select for transduced HeLa cells. Cell colonies were collected and pooled 2 weeks postinfection. Genomic DNA was purified and HTLV-1 integration junction sites were cloned using linker-mediated PCR as described previously (57). Briefly, genomic DNA was digested with the restriction enzyme MseI and ligated to a double-stranded DNA linker. A second enzyme, NheI, was used together with MseI to eliminate amplification of proviral sequences from the long terminal repeat (LTR) at the other end of the virus. Viral integration junction sites were PCR amplified using one primer complementary to the viral LTR and the other primer complementary to the linker. A second round of nested PCR was performed, and the resulting junction site sequences were directly cloned into the TOPO vector (Invitrogen, CA) and sequenced. The HTLV-1 5' LTR primers used in this study include HTLVu5 (5'-GCCGCTACAGATCGAAAGTT-3') and HTLVu5nest (5'-ACGACTAAC TGCCGGCTTG-3'). Linker sequences are Afl3-us (5'-GTAATACGACTCAC TATAGGGCTCCGCTTAAGGGAC-3') and Afl3-ls (5'-PO4-TAGTCCCT TAAG CGAG-NH₂-C7-3'). Primers for linkers are Afl3 (5'-GTAATACGAC TCACTATAGGGC-3') and Afl3nest (5'-AGGGCTCCGCTTAAGGGAC-3').

Analysis of HTLV-1 and other viral integration sites in the human genome. Raw sequences were filtered to select those that had the expected LTR sequence and linker sequences. Sequences were trimmed and aligned to human genome

hg18 (University of California, Santa Cruz [UCSC] March 2006 freeze; NCBI build 36.1) using the Blat program (<http://genome.ucsc.edu>). To be considered an authentic integration site, a clone must meet several criteria: (i) the genome must be matched with >95% identity; (ii) the match must start immediately after the LTR sequence (<5 bp); (iii) the match to the genome must be contiguous with no big gaps; and (iv) if a clone matches multiple genomic sites, the best match is chosen only if it has a Blat score 10 or more higher than the second-best match. With these criteria, we mapped 541 unique HTLV-1 integration sites in the human genome from HeLa cells. Other data sets for HIV, MLV, FV, ASLV, and SIV integration sites were downloaded from GenBank and mapped to the human genome using the same automated program except that a cutoff value of 90% identity was used for SIV integration sites cloned from macaque (25). Customized Perl programs were used to compare localized integration sites to various genomic features. A set of 10,000 random integration sites in the human genome were generated in silico and analyzed together with viral integration sites. All genomic feature tables and chromosome sequences for human genome hg18 were downloaded from the UCSC genome database (<http://genome.ucsc.edu/>). Multiple data sets for each virus were first analyzed separately. We did not observe any statistical difference between subsets, and the data sets for each virus were pooled.

Cluster analysis of viral integration site profiles and phylogenetic analysis of viral integrase homology. BRB-arrayTools 3.3.0 software (<http://linus.nci.nih.gov/BRB-ArrayTools.html>) was used to cluster viral integration site profiles. Integration sites for all six retroviruses and random sites were analyzed using a total of 69 genomic features, including genes, CpG islands, GC content, etc. Unsupervised hierarchical clustering was performed using 69 genomic features with Euclidean distance and average linkage.

For phylogenetic analysis, amino acid sequences of viral integrase for all six retroviruses were aligned with the AlignX program based on the Clustal W algorithm in the VectorNTI software suite (Invitrogen). The SwissProt accession numbers are as follows: P14078 (HTLV-1), Q7SQ98 (ASLV), P23074 (FV), P03355 (MLV), P05896 (SIV), and P03366 (HIV). Reverse transcriptase and RNase H sequences were trimmed off. Only the integrase sequences of the POL proteins were used for alignment. An unrooted neighbor-joining tree was generated with Mega3.1 software with 10,000 bootstrap samples (29). A phylogenetic tree was also generated by the GeneBee TreeTop phylogenetic tree prediction server based on a cluster algorithm (http://www.genebee.msu.su/services/phtree_reduced.html) (5). The two trees were very similar.

RESULTS

Cloning and mapping of HTLV-1 integration sites in the human genome. Infection with wild-type HTLV-1 poses several problems for the analysis of integration sites, which include low infectious titers and deleterious effects of viral gene products. HTLV-1 gene expression can induce either cell proliferation or cell death, depending on the target cell, and chronically infected cells will be either positively or negatively selected for virus expression. We circumvented these problems by using HTLV-1 vectors that do not express viral proteins (18) but do express a selectable marker. In order to develop a large set of HTLV-1 integrations, 293T effector cells, which produce HTLV-1 virus-like particles encoding a neomycin resistance gene, were cocultured with HeLa target cells. 293T effector cells were treated with mitomycin prior to the cocultivation. Infected HeLa cells were selected in G418 to enrich for cells with integration events. More than 3,000 colonies were collected and pooled 2 weeks postinfection. Genomic DNA was purified, and HTLV-1 integration junction sites were cloned using linker-mediated PCR as described previously (57). Sequences were trimmed and aligned to human genome hg18 (UCSC March 2006 freeze; NCBI Build 36.1) using the Blat program (<http://genome.ucsc.edu>). We cloned and mapped 541 unique HTLV-1 integration sites from the HeLa cells. Published data sets for other retroviruses were also mapped to human genome hg18 and compared to each other (Table 1). For each virus for which multiple data sets were available, we

TABLE 1. Integration site data sets used in this study

Data set and no. of integration sites	Cell type(s)	Reference	GenBank accession no.
HTLV-1; 541	HeLa	This report	EF580177–EF580913
ASLV; 606	293T-TVA	Mitchell et al. (40)	CL528303–CL528772
	HeLa	Narezkina et al. (41)	AY653309–AY653534
FV; 3,293	Fibroblast, CD34 ⁺ cells	Trobridge et al. (51)	DU798511–DU799518, DU796690–DU798510
	293	Nowrouzi et al. (42)	DQ193477–DQ193515
MLV; 1,383	HeLa	Wu et al. (57), Lewinski et al. (34)	AY515855–AY516880, DX598305–DX598906
SIV; 378	CEMx174	Crise et al. (17)	AY679815–AY680027
	Rhesus macaque CD34 ⁺ cell	Hematti et al. (25)	AY728482–AY728804
HIV; 2,636	PBMC, IMR-90	Mitchell et al. (40)	CL529240–CL529767, CL528773–CL529239
	SupT1	Schroder et al. (49)	BH610086–BH609398
	H9, HeLa	Wu et al. (57)	AY516881–AY517469
	CD34 ⁺ cell	Trobridge et al. (51)	DU799519–DU800849

analyzed each data set for the virus separately and found no significant differences. The multiple data sets for the individual viruses were pooled and used for most analyses reported here.

HTLV-1 integration target sites exhibit a palindromic consensus at the integration site. When retroviruses integrate into the host genome, a small host target site sequence (4 to 6 bp) is duplicated at both ends of the viral DNA. The proviruses of lentiviruses, including HIV and SIV, generate 5-bp target site duplications. MLV and FV generate 4-bp target site duplications. ASLV and HTLV-1 generate 6-bp target site duplications, although ASLV also generates 5-bp target site duplications in approximately 25% of its proviral insertions (43). Previous

analysis of local DNA sequences around the target sites has shown that a palindromic consensus sequence is a common feature for the proviruses of HIV, SIV, MLV, and ASLV (26, 58). We analyzed the genomic sequences upstream and downstream of the HTLV-1 integration sites by aligning all sequences relative to the integration site (position 1), in the same orientation relative to the provirus. The nucleotide frequency at each position was compared to the expected value for random sites, which in the human genome is about 30% A, 30% T, 20% G, and 20% C. As with other proviruses, the sequences around HTLV-1 integration sites also showed a palindromic consensus centered on the target site duplication (Fig. 1). At



FIG. 1. Palindromic consensus sequences at retroviral integration sites. Base compositions around the integration sites were calculated. Integration occurs between positions -1 and 1 on the top strand. Colored positions have frequencies of bases statistically different from those of randomly generated positions ($P < 0.01$), which are 30%/20%/20%/30% for A/C/G/T. Bases with a greater than 10% increase of frequency at a position are colored green, and bases with a greater than 10% decrease of frequency at a position are colored red. Inferred duplicated target sites are in the blue box, and DNA strand transfer occurs at positions labeled by arrows. The base preferences show palindromic patterns centered on the duplicated target sites, and the symmetries are marked with the dotted vertical line.

TABLE 2. Integration frequency near genomic features

Genomic feature	Frequency (%) for random sites	Frequency (%) of integration near genomic feature					
		HTLV-1	ASLV	FV	MLV	SIV	HIV
Within ± 2 kb of a TxStart site of a RefSeq gene	2.5	5.2 ^a	3.8	10.4 ^a	18.0 ^a	0.8	1.9
Within ± 2 kb of a CpG island	4.3	5.9	7.6 ^a	15.2 ^a	21.5 ^a	1.3	3.6
Within ± 2 kb of a FirstEF promoter	4.8	6.8	8.3 ^a	16.4 ^a	23.7 ^a	1.3	3.6
Within ± 2 kb of a FirstEF first exon	4.8	6.7	8.4 ^a	16.1 ^a	23.7 ^a	2.1	3.9
Within ± 2 kb of a FirstEF CpG window	4.0	5.7	6.1	14.5 ^a	21.1 ^a	0.5 ^a	2.5 ^a
Within ± 1 kb of DNase-hypersensitive sites (score, ≥ 750)	0.37	0.55	0.99	2.25 ^a	6.00 ^a	0.53	0.11
Within RefSeq genes	35.7	46.8 ^a	46.4 ^a	32.7	45.7 ^a	80.2 ^a	72.2 ^a
Within Known genes	35.2	46.8 ^a	45.5 ^a	32.0 ^a	44.0 ^a	74.9 ^a	70.6 ^a
Within Ensembl genes	34.2	46.6 ^a	45.4 ^a	31.7	44.1 ^a	70.9 ^a	68.3 ^a
Within Genscan genes	69.8	68.8	70.0	66.5 ^a	65.4 ^a	81.5 ^a	77.5 ^a
Gene density within 1 Mb	17.8	21.0 ^b	20.6 ^b	20.8 ^b	27.7 ^b	36.7 ^b	31.1 ^b
GC content, 100 bp	40.8	39.5	40.7	42.9 ^b	42.6 ^b	36.9 ^b	38.4 ^b

^a $P < 0.0001$, chi-square test, comparison to value for random-site sequences.

^b $P < 0.001$, t test, comparison to value for random-site sequences.

position 1, nucleotide T was disfavored, whereas A was disfavored at position 6 within the duplication. Outside the duplication at position -2 , T was strongly favored and G was disfavored. This is reflected in the symmetrical position $+2$, where A was favored and C was disfavored. The data are consistent with what was reported earlier for HTLV-1 integration sites in patients (31).

Integration near transcription start site. Based on the analysis of multiple retroviral integration sites, it has been shown that integration is more or less likely to occur near certain features within the genome (40, 49, 57). In particular, MLV integration preferentially occurs near transcription start sites or promoter regions (57). The frequency of HTLV integration near transcription start sites or promoters was compared with those for 5 other retroviruses and with 10,000 computer-generated random sites. Although transcription start sites and promoter regions have not been completely annotated in the human genome, we used several tables in the UCSC genome database to estimate the proximities of integration sites to these features. First, we looked at the frequency of integration near the transcription start sites of RefSeq genes (Table 2 and Fig. 2A). MLV, as previously reported, showed the strongest preference for integration near transcription start sites, with 18.0% of the integration sites within a ± 2 -kb window of transcription start sites ($P < 0.0001$, compared to random sites). FV showed the second-strongest preference for integration near transcription start sites, with 10.4% of the integration sites within the same window ($P < 0.0001$, compared to random sites). HTLV showed a weak but significant preference for integration near transcription start sites, with 5.2% of integrations within this window ($P = 0.0002$). ASLV, SIV, and HIV showed no significant preference compared to random sites (3.8%, 0.8%, 1.9%, and 2.5%, respectively).

For MLV and FV, the frequency of integration sites is bell shaped relative to transcription start sites, i.e., the frequency of integration is higher nearer transcriptional start sites. FV shows a small shift of peak position toward the regions upstream of transcriptional start sites (Fig. 2A). For SIV and HIV, there was a small reduction in the frequency of integration near (within 1 kb) transcription start sites ($P = 0.09$ for SIV; $P = 0.009$ for HIV).

Integration near CpG islands. CpG islands are thought to be associated with transcriptional start sites in vertebrate genomes (3, 30). We analyzed integration sites of all six retroviruses relative to the random data set for proximity to CpG islands in the human genome (Table 2; Fig. 2B). Again, MLV showed the strongest preference for integration into regions near CpG islands, with 21.5% of integration sites within a ± 2 -kb window of CpG islands ($P < 0.0001$). FV showed the second-strongest preference near CpG islands, with 15.2% of integration sites within the same window ($P < 0.0001$). ASLV showed a slight preference for regions around CpG islands, with 7.6% ($P = 0.0001$) of its integration sites within the window. HTLV and HIV showed no significant preference compared to that for random sites (5.9%, 3.6%, and 4.3%, respectively). The frequency of SIV integrations near CpG islands (1.3%) was lower than that for random sites, although not statistically significant. For each of the viruses, the integration frequency near CpG islands is in good agreement with the frequency near transcription start sites.

In addition, we used the FirstEF (First Exon Finder) table from the UCSC genome database to estimate the integration frequency near transcription start sites or promoter regions. FirstEF is a program that predicts promoters and 5'-terminal exons. The FirstEF database contains three types of predictions for the human genome: first exon, promoter, and CpG window. The integration frequencies relative to these three features were similar to the data from the RefSeq transcription start sites and the CpG islands for each of the viruses. For MLV, 23.7% of the integration sites were within the ± 2 -kb window of predicted promoters ($P < 0.0001$). FV integrated in the same regions at a frequency of 16.4% ($P < 0.0001$). HTLV and ASLV showed a weak preference for promoter regions (6.8% [$P = 0.03$] and 8.3% [$P = 0.0001$], respectively). HIV and SIV showed no preference or a slight avoidance for these regions compared to random sites (HIV, 3.6%; SIV, 1.3%; and random, 4.8%, respectively).

Integration in genes. HIV and SIV were reported to preferentially integrate into genes or transcription units (17, 25, 49, 57). The frequency of HTLV integration into genes or transcription units was compared to those for five other viruses and a random data set (Table 2). Several human gene annotation

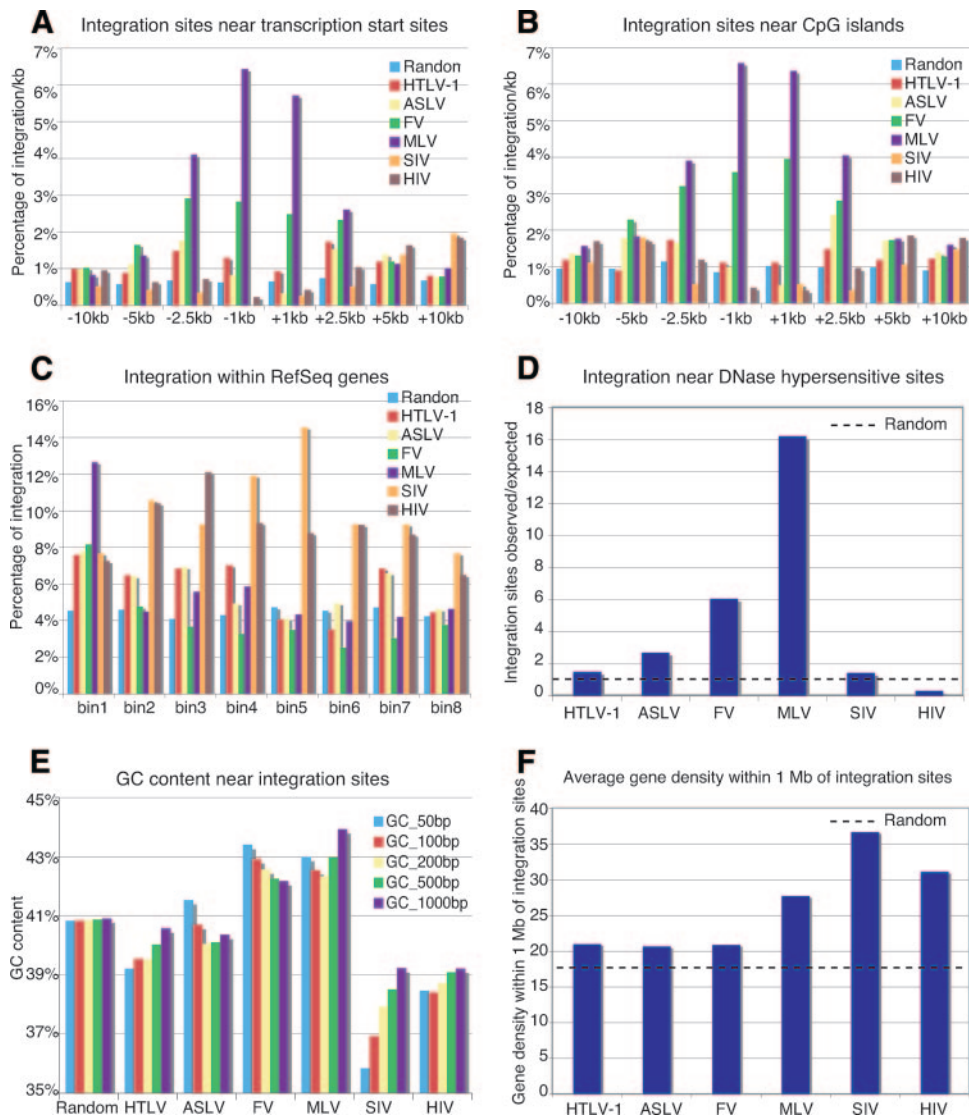


FIG. 2. Integration frequencies of HTLV-1 and five other retroviruses near various genomic features. (A) Integration frequency near transcription start sites of RefSeq genes. The frequency is shown as the percentage of integration sites adjusted to the density as numbers per kb for each interval near the transcription start sites of RefSeq genes. $-$, denotes region upstream of transcription start sites; $+$, region downstream of transcription start sites. (B) Integration frequency near CpG islands. The percentage of the integration sites (per kb) is shown for each interval near CpG islands. $-$, region upstream of CpG islands; $+$, region downstream of CpG islands. (C) Integration site distribution within RefSeq genes. Each gene was conceptually divided into eight equal-size bins from the start to the end of the gene. Integration sites in each bin were added together and plotted for all viruses. (D) Integration frequencies near DNase hypersensitive sites. Integration sites within ± 1 kb of DNase I cleavage sites were compared to random expected values. Frequency is represented as the ratio of observed sites/expected sites. Frequency near random sites is represented by the dotted line. (E) GC content near integration sites. The GC content within various sizes of windows near integration sites were computed for each virus and compared to random integration sites. The GC content near random sites reflects that the average GC content for the human genome is close to 41% GC. The GC contents around MLV and FV integration sites are higher than those for random sites, while the GC contents around SIV and HIV integration sites are lower than those for random sites. (F) Gene density around integration sites. RefSeq genes found within ± 1 Mb of integration sites were averaged for each virus and compared to 10,000 random sites. The dotted line represents gene density around random sites.

tables from the UCSC genome database were used for this analysis, including RefSeq genes, Known genes, Ensembl genes, MGC genes, SGP genes, and Genescan genes. We found that regardless of which database was used, a consistent pattern was seen for each virus except in the case of Genescan genes, which are totally computationally predicted. Here we focus on RefSeq genes, because they are well annotated. Our analysis of ASLV, FV, MLV, SIV, and HIV agrees with the published reports. The SIV and HIV proviruses were preferentially integrated

into genes, with 80% and 72% of integration sites, within RefSeq genes ($P < 0.0001$). HTLV, like ASLV and MLV, showed a modest preference for genes, with a ratio of 46.8%, 46.4%, and 45.7%, respectively, in RefSeq genes ($P < 0.0001$). FV showed no preference for genes; only 32.7% of FV integrations were within RefSeq genes, even lower than the random data set, which has 35.7% within RefSeq genes, suggesting FV may avoid genes as targets ($P = 0.002$).

We also looked at the distribution of integration sites within

RefSeq genes (Fig. 2C). All genes are divided into eight bins, starting from the transcription start site. Integration sites inside genes were placed in those eight bins according to location. The percentage of integration sites was then calculated for each bin. For MLV and FV, the first bin has the highest integration frequency ($P < 0.05$), reflecting their preference for transcription start sites. For SIV and HIV, the frequency tends to be higher in the middle of genes (second to seventh bins) and lower at both ends of genes (first and eighth bins). HTLV and ASLV showed a roughly even distribution across all eight bins.

Integration near DNase-hypersensitive sites. DNase-hypersensitive sites are believed to be nucleosome-free regions of the chromatin associated with regulatory elements, such as promoters, silencers, enhancers, and locus control regions in the genome (21). Recently Crawford et al. mapped a large number of DNase I-hypersensitive sites in the human genome (15, 16). DNase-hypersensitive sites were enriched upstream of genes, in CpG islands, and in regions that are conserved in multiple species. Most of the DNase-hypersensitive sites were not cell line specific. Figure 2D shows the integration preferences of all six retroviruses within a ± 1 -kb window of all DNase-hypersensitive sites with a score of 750 (this score correlates with approximately 85% of the valid DNase-hypersensitive sites; NHGRI DNase I-hypersensitive sites track description, <http://genome.ucsc.edu/>). Among the six retroviruses, MLV showed the strongest preference for integrating near DNase-hypersensitive sites ($P < 0.0001$), while FV showed a weaker yet still significant preference for DNase-hypersensitive sites ($P < 0.0001$). HTLV, ASLV, SIV, and HIV showed no significant preference for DNase-hypersensitive sites compared to random sites.

GC content near integration sites. Genomic sequences around integration sites were aligned, and GC content in variously sized windows (50 bp, 100 bp, 200 bp, 500 bp, and 1,000 bp) was computed. Table 2 and Fig. 2E show the average GC content in these windows around the integration sites of all six retroviruses. MLV and FV integration sites have a higher GC content than the random sites in window sizes up to 1 kb ($P < 0.0001$, Monte Carlo simulation, compared to $100,000 \times n$ sets of random sites, where n is the matched number of integration sites used for each virus). These results may reflect the preferences for CpG islands by MLV and FV. SIV and HIV both have lower GC content surrounding integration sites than for random sites ($P < 0.0001$). The GC content surrounding HTLV-1 and ASLV sites was similar to that for random sites.

Integration and gene density. The gene densities surrounding the integration sites of all six retroviruses were also calculated. The average number of genes found within 1 Mb of the integration sites (Table 2) for each virus was plotted and is shown in Fig. 2F. All viruses showed an elevated average gene density within a 1-Mb window of the integration sites ($P < 0.0001$, compared to 10,000 random sites with a t test). The highest gene density was found around SIV integration sites. HIV integration sites had the second-highest gene density. MLV integration sites had the third-highest gene density. Gene densities around HTLV-1, ASLV, and FV integration sites were similar.

Global comparison of integration target site preferences of six retroviruses. From the above analysis of integration sites of

six retroviruses, it appeared that HTLV-1 and ASLV integration sites were similar with respect to the integration preferences for genomic features such as transcription start sites, CpG islands, promoters, DNase-hypersensitive sites, genes, gene density, and GC content. So were FV and MLV integration sites, as well as SIV and HIV integration sites. Clustering methods have been commonly used to measure the similarities and differences within and between groups of samples. A machine learning algorithm was recently used by Lewinski et al. to describe the similarity of global integration profiles of HIV, MLV, and HIV/MLV hybrid viruses (34). We performed cluster analysis of the global integration profiles of six retroviruses and the random-site control. This was done by taking into account 69 different genomic features, some of which have been described above (see Table S1 in the supplemental material). Using unsupervised hierarchical clustering, with euclidean distance and average linkage, six viruses and the random sites could be clearly separated into three distinct clusters (Fig. 3). SIV and HIV form one cluster. FV and MLV form a second cluster, while HTLV-1 and ASLV form a cluster with the random sites.

To ask whether the clustering of retroviruses based on integration profiles could be correlated with a common genetic trait of the viruses, we performed phylogenetic analysis based on the integrases encoded by the six viruses. Our results show that these six retroviruses can be grouped into the same three clusters based on the amino acid sequence similarity of their integrases (Fig. 3), and this phylogenetic grouping is in good agreement with previous analysis of the relatedness of the viruses (20).

DISCUSSION

One of the technical difficulties in working with HTLV is to produce the high-titer cell-free virus stocks necessary to generate large numbers of independent provirus integration sites. The alternative, using chronically infected cells, could lead to biases from the selective pressure imposed by the effects of integration and/or virus gene products on cell growth. We used an HTLV-1 vector system that makes it possible to select for infected cells without the expression of virus genes in the target cells. This raises a question of whether the integration sites we recovered accurately reflect the complete set of integration sites. We believe that the effect of drug selection will be modest, both because the time of selection is relatively short (2 weeks) and because previous studies have demonstrated that a short-term drug selection did not significantly affect the populations of recovered integration sites (34). However, as discussed below, we cannot exclude the possibility that cell type might influence integration site profiles for HTLV-1.

We observed two hot spots (chr11p11.2 and chr11q12.1) for HTLV-1 integration in HeLa cells. The first, on chr11p11.2, is a 162-kb region that had 6 independent integration sites ($P = 0.00001$ based on $100,000 \times 541$ Monte Carlo simulations). This is the location for the 5' end of the gene encoding a receptor protein, tyrosine phosphatase J. Tyrosine phosphatase J is present in all hematopoietic lineages and was shown to negatively regulate T-cell receptor signaling (1). The second hot spot is a 100-kb region on chr11q12.1 that had 5 independent integration sites ($P = 0.0004$, based on $100,000 \times 541$

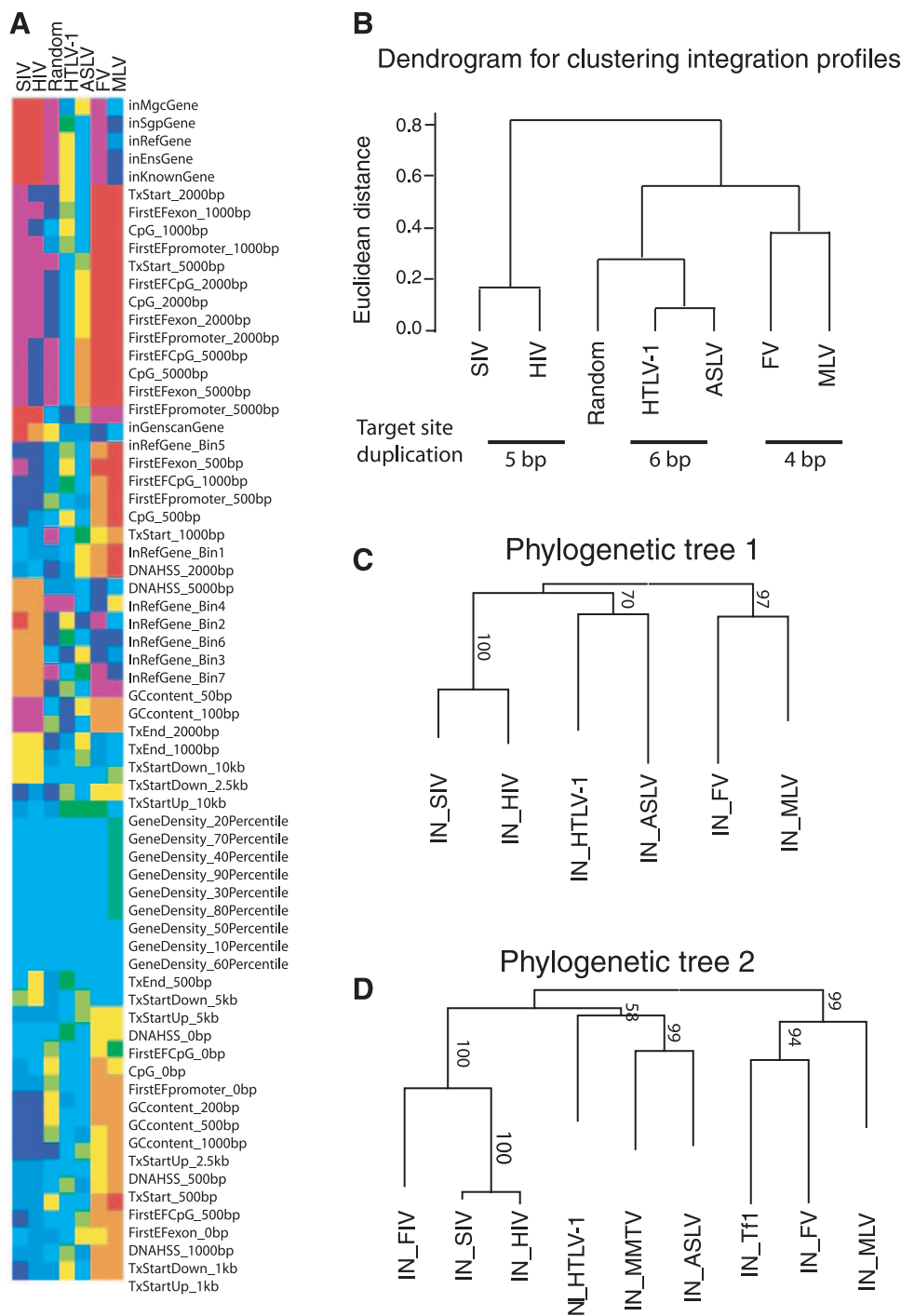


FIG. 3. Clustering of integration site preferences and phylogenetic analysis of integrases of all six retroviruses. (A) Heat map of clustering of the integration sites for all 6 retroviruses and random sites based on 69 genomic features. (B) Dendrogram based on location of integration in relation to 69 genomic features. Unsupervised hierarchical clustering, with euclidean distance and average linkage was used to generate the dendrogram. (C) Phylogenetic tree based on amino acid sequences of the six retroviral integrases. Bootstrap values for the neighbor-joining method (percentages from 10,000 trials) are shown on each branch. Integrase sequences are from the following POL proteins: P14078 (HTLV-1), Q7SQ98 (ASLV), P23074 (FV), P03355 (MLV), P05896 (SIV), and P03366 (HIV). (D) Phylogenetic tree with additional integrase sequences, including P31822 (FIV), P03365 (MMTV), and AAA35339 (Tf1) integrases, showing the three integrases been placed into different clusters.

Monte Carlo simulations). There are two genes within this region: RTN4RL2 and SLC43A1. We do not know the biological relevance of the hot spots or whether the hot spots were related to the drug selection. Earlier work with HIV also found

an integration hot spot in SupT1 cells, but this hot spot did not appear in other cell types studied (49). Our results show that HTLV-1 integration is nearly random within the HeLa cell genome. The six retroviruses compared

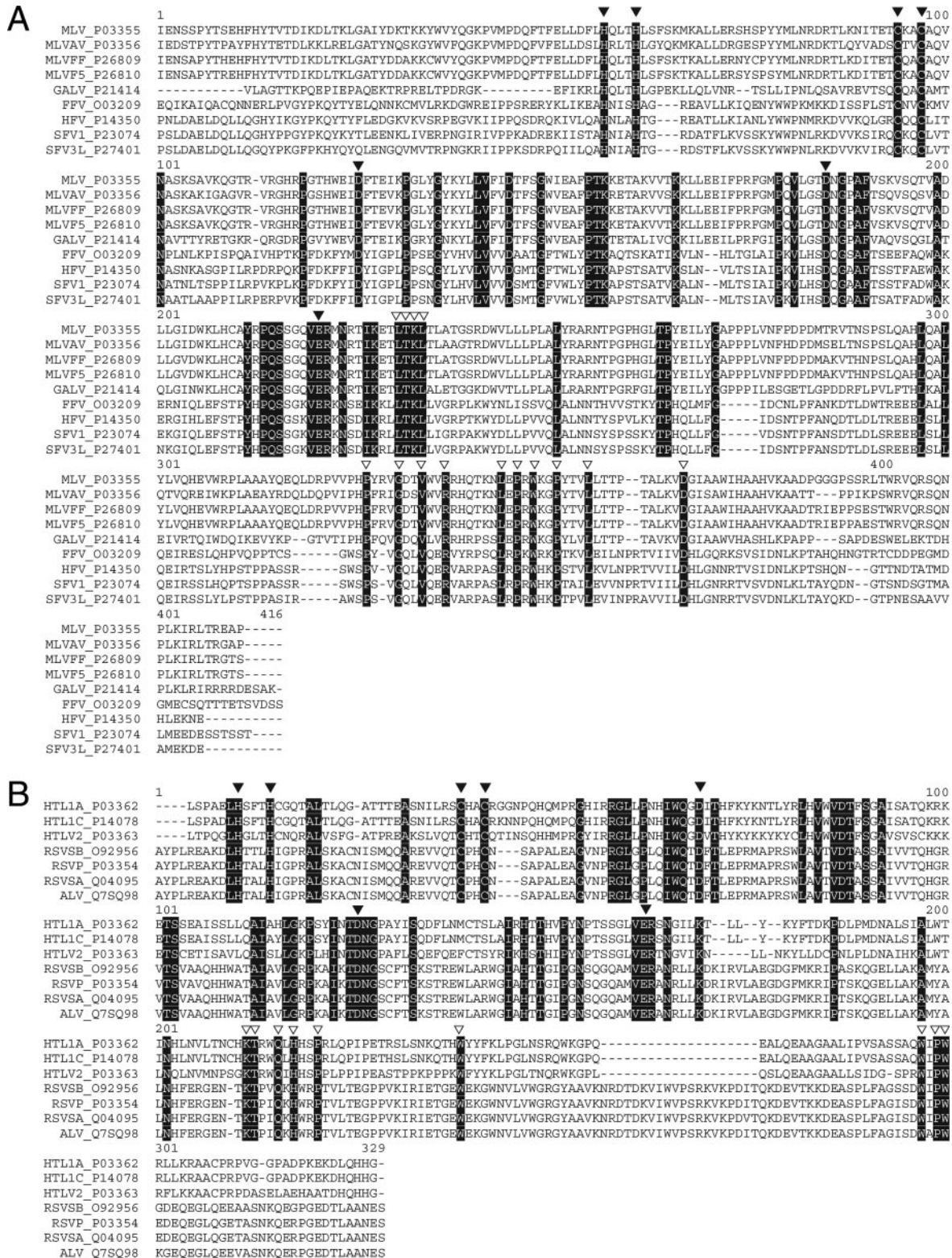


FIG. 4. Alignment of retroviral integrases within each cluster reveals conserved motifs outside the catalytic core that may interact with cellular targeting factors. Identical amino acids are labeled with a black background. The zinc finger motif (HHCC) and the catalytic core (DDE motif) are labeled with black arrowheads. Conserved residues of other motifs are labeled with white arrowheads. (A) Alignment of IN from the MLV and FV families revealed additional conserved motifs. The LTKL motif is probably part of the catalytic core, based on a domain structure study comparing HIV and MLV IN (48). Another conserved motif (PxxxGxxVxxRxxxxxxLxP(R/K)xxPxxxL) is found in the C-terminal regions of MLV and FV IN. (B) Alignment of IN from the HTLV and ALV families. The KTxQxHxxP motif is located on the linker between the catalytic core and the C-terminal domain based on the RSV crystal structure (1C0M). Another conserved motif, WxPW, is found at the ends of the C-terminal regions of HTLV and ALV IN.

here can be placed into three groups, based on the preferences of their integration sites for different genomic features. The groups are characterized by integration sites that are predominantly as follows: (i) near transcription start sites and CpG islands (MLV and FV); (ii) within genes or transcription units (SIV and HIV); or (iii) randomly dispersed (HTLV and ASLV). The same three pairs of retroviruses were clustered together in phylogenetic analyses of their integrase proteins, even though viruses in two of these pairs were from different retroviral genera. These results suggest that the most closely related integrase proteins direct integration into regions of the genome with similar features and that viruses in these different groups use distinct mechanisms to access their integration sites.

It should be possible to predict the global integration profiles of uncharacterized retroviruses based on integrase phylogenies. For example, feline immunodeficiency virus (FIV), which is being used to develop gene therapy vectors (45), has a 5-bp target duplication site. Phylogenetic analysis puts FIV integrase in the same cluster with SIV and HIV, predicting that the integration profile will be similar to that of SIV or HIV. The recent report by Kang et al. on FIV vector integration sites is consistent with this prediction (27). The relationship of integrase phylogeny and integration preference can be extended to certain retrotransposons. The Tf1 transposon from *Schizosaccharomyces pombe* is an LTR retrotransposon closely related to retroviruses (33). Phylogenetic analysis of Tf1 integrase places it in the MLV/FV cluster (Fig. 3). It has been shown that the Tf1 integration site preference resembles MLV/FV in that Tf1 prefers to integrate in the promoter regions of polymerase II-transcribed genes (4, 50). The integration profile of mouse mammary tumor virus (MMTV), a betaretrovirus that generates a 6-bp target site duplication, has not been determined. Phylogenetic analysis of MMTV integrase (Fig. 3) places it in the HTLV and ASLV cluster, leading to the prediction that MMTV will integrate into the host genome with little preference for any genomic features we have analyzed.

Both cellular and viral factors may contribute to the integration sites selected by retroviruses (6, 7, 56). Cellular factors can cooperate in the targeting of preintegration complexes to specific genomic features (8). LEDGF/p75 binds to HIV integrase (11, 35, 37, 53), increases the efficiency of HIV integration (36), and plays a role in targeting integration into genes (12). In contrast, MLV integrase does not interact with LEDGF/p75 but is likely to target promoter regions by interacting with different cellular factors. The absence of integration site specificity for HTLV-1 and ASLV could be due to interactions with ubiquitous chromosomal proteins or to a lack of interaction with host proteins. Alternatively, we cannot rule out the possibility that the cellular protein or protein isoform that interacts with HTLV-1 or ASLV integrase is not expressed in HeLa cells. Further studies of integration profiles for these viruses in other cell types will be needed to resolve these issues.

Although the interaction between retroviral integrases and the host factors involves the three-dimensional structure of the proteins, as illustrated by lentiviral integrase and LEDGF/p75 (10), alignment of primary sequences of related proteins often reveals important motifs. To identify potential interaction motifs that are shared among integrase proteins, integrase sequences from retroviruses with similar integration site prefer-

ences were aligned. Apparent conserved regions were observed (Fig. 4). For instance, alignment of MLV, FV, and other closely related integrases revealed conserved motifs in addition to the HHCC zinc finger motif and the DDE catalytic motif (Fig. 4A). The LTKL motif is probably within the $\alpha 4$ helix of the catalytic domain, based on the comparison of domain structures of MLV and HIV IN (48). Further toward the C terminus, another conserved region can be defined as GxxVxxRxxxxxxLxP(R/K)WxxPxx(V/I)L, where x is any amino acid. This domain was also identified as a conserved domain (the GPY/F domain) in the Ty3/Gypsy class of LTR retrotransposons and some retroviral integrases (38), although the element we have identified varies slightly from the reported GPY/F module. This motif in the Ty3/Gypsy class of retrotransposons was proposed to play a role in directing integration specificity (38). This domain is also present in the *Schizosaccharomyces pombe* Tf1 element, which has an integration site preference similar to those of MLV and FV, targeting upstream regions of polymerase II-transcribed genes (4, 50). This domain is not found in other retroviral integrases analyzed in this study. Conserved motifs were also observed when HTLV-1 and ASLV families were aligned (Fig. 4B). It will be interesting to see if mutations in these regions alter the targeting specificities of the integrases.

The studies presented here also have implications for the development of gene therapy vectors. The risk of insertional mutagenesis by retroviral vectors is exemplified by the development of leukemia from vector DNA insertion in the X-SCID gene therapy trial (9, 22, 23). Our results show that there is no correlation between integration site profiles and the potential for insertional mutagenesis. For example, HTLV-1 and ASLV have very similar integration preferences, but whereas ASLV infection is notorious for generating transformed cells via proviral insertions near proto-oncogenes, cellular transformation by HTLV-1 is more likely due to the effects of the expression of the viral protein Tax than to the site of insertion. MLV and FV have similar integration profiles, but FV infections have not been associated with tumor formation, while MLV infections frequently cause tumor formation by viral DNA insertions. HIV and SIV have a strong preference for targeting genes, but there have been no reports of insertional mutagenesis during the course of HIV or SIV infection despite high levels of virus replication in the host. Rather than correlating with integration site targeting, it is more likely that cell transformation by insertional mutagenesis is related to cell tropism, levels of infectious spread within the host, and the transcriptional activity of the viral promoter. It is noteworthy that MLV and ASLV have relatively strong constitutive promoters, in contrast to HTLV-1, FV, and HIV/SIV, which are complex retroviruses that encode transacting proteins which control transcription and RNA transport. It is clear that factors other than integration site preference make strong contributions to the risks for different retroviral vectors, and all the causes need to be addressed for vectors intended for use in human gene therapy.

ACKNOWLEDGMENTS

This project was funded in whole or in part by federal funds from the National Cancer Institute, National Institutes of Health, under contract no. N01-CO-12400.

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

REFERENCES

- Baker, J. E., R. Majeti, S. G. Tangye, and A. Weiss. 2001. Protein tyrosine phosphatase CD148-mediated inhibition of T-cell receptor signal transduction is associated with reduced LAT and phospholipase C γ 1 phosphorylation. *Mol. Cell. Biol.* 21:2393–2403.
- Barr, S. D., A. Ciuffi, J. Leipzig, P. Shinn, J. R. Ecker, and F. D. Bushman. 2006. HIV integration site selection: targeting in macrophages and the effects of different routes of viral entry. *Mol. Ther.* 14:218–225.
- Bird, A. P. 1986. CpG-rich islands and the function of DNA methylation. *Nature* 321:209–213.
- Bowen, N. J., I. K. Jordan, J. A. Epstein, V. Wood, and H. L. Levin. 2003. Retrotransposons and their recognition of pol II promoters: a comprehensive survey of the transposable elements from the complete genome sequence of *Schizosaccharomyces pombe*. *Genome Res.* 13:1984–1997.
- Brodski, L. I., V. V. Ivanov, I. L. Kalaidzidis, A. M. Leontovich, V. K. Nikolaev, S. I. Feranchuk, and V. A. Drachev. 1995. GeneBee-NET: an Internet based server for biopolymer structure analysis. *Biokhimiia* 60:1221–1230. (In Russian.)
- Bushman, F. 2002. Targeting retroviral integration? *Mol. Ther.* 6:570–571.
- Bushman, F., M. Lewinski, A. Ciuffi, S. Barr, J. Leipzig, S. Hannenhalli, and C. Hoffmann. 2005. Genome-wide analysis of retroviral DNA integration. *Nat. Rev. Microbiol.* 3:848–858.
- Bushman, F. D. 2003. Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. *Cell* 115:135–138.
- Check, E. 2005. Gene therapy put on hold as third child develops cancer. *Nature* 433:561.
- Cherepanov, P. 2007. LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res.* 35:113–124.
- Cherepanov, P., G. Maertens, P. Proost, B. Devreese, J. Van Beeumen, Y. Engelborghs, E. De Clercq, and Z. Debyser. 2003. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* 278:372–381.
- Ciuffi, A., M. Llano, E. Poeschla, C. Hoffmann, J. Leipzig, P. Shinn, J. R. Ecker, and F. D. Bushman. 2005. A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.* 11:1287–1289.
- Ciuffi, A., R. S. Mitchell, C. Hoffmann, J. Leipzig, P. Shinn, J. R. Ecker, and F. D. Bushman. 2006. Integration site selection by HIV-based vectors in dividing and growth-arrested IMR-90 lung fibroblasts. *Mol. Ther.* 13:366–373.
- Coffin, J. M., S. H. Hughes, and H. E. Verma. 1997. *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Crawford, G. E., I. E. Holt, J. C. Mullikin, D. Tai, R. Blakesley, G. Bouffard, A. Young, C. Masiello, E. D. Green, T. G. Wolfsberg, and F. S. Collins. 2004. Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc. Natl. Acad. Sci. USA* 101:992–997.
- Crawford, G. E., I. E. Holt, J. Whittle, B. D. Webb, D. Tai, S. Davis, E. H. Margulies, Y. Chen, J. A. Bernat, D. Ginsburg, D. Zhou, S. Luo, T. J. Vasicsek, M. J. Daly, T. G. Wolfsberg, and F. S. Collins. 2006. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res.* 16:123–131.
- Crise, B., Y. Li, C. Yuan, D. R. Morcock, D. Whithy, D. J. Munroe, L. O. Arthur, and X. Wu. 2005. Simian immunodeficiency virus integration preference is similar to that of human immunodeficiency virus type 1. *J. Virol.* 79:12199–12204.
- Derse, D., S. A. Hill, P. A. Lloyd, H. Chung, and B. A. Morse. 2001. Examining human T-lymphotropic virus type 1 infection and replication by cell-free infection with recombinant virus vectors. *J. Virol.* 75:8461–8468.
- Doi, K., X. Wu, Y. Taniguchi, J. Yasunaga, Y. Satou, A. Okayama, K. Nosaka, and M. Matsuoka. 2005. Preferential selection of human T-cell leukemia virus type I provirus integration sites in leukemic versus carrier states. *Blood* 106:1048–1053.
- Doolittle, R. F., D. F. Feng, M. A. McClure, and M. S. Johnson. 1990. Retrovirus phylogeny and evolution. *Curr. Top. Microbiol. Immunol.* 157: 1–18.
- Gross, D. S., and W. T. Garrard. 1988. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* 57:159–197.
- Hacein-Bey-Abina, S., C. von Kalle, M. Schmidt, F. Le Deist, N. Wulfraat, E. McIntyre, I. Radford, J. L. Villeval, C. C. Fraser, M. Cavazzana-Calvo, and A. Fischer. 2003. A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* 348:255–256.
- Hacein-Bey-Abina, S., C. Von Kalle, M. Schmidt, M. P. McCormack, N. Wulfraat, P. Leboulch, A. Lim, C. S. Osborne, R. Pawliuk, E. Morillon, R. Sorensen, A. Forster, P. Fraser, J. I. Cohen, G. de Saint Basile, I. Alexander, U. Wintergerst, T. Frebourg, A. Aurias, D. Stoppa-Lyonnet, S. Romana, I. Radford-Weiss, F. Gross, F. Valensi, E. Delabesse, E. Macintyre, F. Sigaux, J. Soulier, L. E. Leiva, M. Wissler, C. Prinz, T. H. Rabbitts, F. Le Deist, A. Fischer, and M. Cavazzana-Calvo. 2003. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302: 415–419.
- Hanai, S., T. Nitta, M. Shoda, M. Tanaka, N. Iso, I. Mizoguchi, S. Yashiki, S. Sonoda, Y. Hasegawa, T. Nagasawa, and M. Miwa. 2004. Integration of human T-cell leukemia virus type 1 in genes of leukemia cells of patients with adult T-cell leukemia. *Cancer Sci.* 95:306–310.
- Hematti, P., B. K. Hong, C. Ferguson, R. Adler, H. Hanawa, S. Sellers, I. E. Holt, C. E. Eckfeldt, Y. Sharma, M. Schmidt, C. von Kalle, D. A. Persons, E. M. Billings, C. M. Verfaillie, A. W. Nienhuis, T. G. Wolfsberg, C. E. Dunbar, and B. Calmels. 2004. Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. *PLoS Biol.* 2:e423.
- Holman, A. G., and J. M. Coffin. 2005. Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proc. Natl. Acad. Sci. USA* 102:6103–6107.
- Kang, Y., C. J. Moretti, T. E. Scheetz, L. Xie, D. T. Tran, T. L. Casavant, P. Ak, C. J. Benham, B. L. Davidson, and P. B. McCray, Jr. 2006. Integration site choice of a feline immunodeficiency virus vector. *J. Virol.* 80:8820–8823.
- Kitamura, Y., Y. M. Lee, and J. M. Coffin. 1992. Nonrandom integration of retroviral DNA in vitro: effect of CpG methylation. *Proc. Natl. Acad. Sci. USA* 89:5532–5536.
- Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* 5:150–163.
- Larsen, F., G. Gundersen, R. Lopez, and H. Prydz. 1992. CpG islands as gene markers in the human genome. *Genomics* 13:1095–1107.
- Leclercq, I., F. Mortreux, M. Cavois, A. Leroy, A. Gessain, S. Wain-Hobson, and E. Wattel. 2000. Host sequences flanking the human T-cell leukemia virus type 1 provirus in vivo. *J. Virol.* 74:2305–2312.
- Leclercq, I., F. Mortreux, A. S. Gabet, C. B. Jonsson, and E. Wattel. 2000. Basis of HTLV type 1 target site selection. *AIDS Res. Hum. Retrovir.* 16:1653–1659.
- Levin, H. L., D. C. Weaver, and J. D. Boeke. 1990. Two related families of retrotransposons from *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* 10:6791–6798.
- Lewinski, M. K., M. Yamashita, M. Eberman, A. Ciuffi, H. Marshall, G. Crawford, F. Collins, P. Shinn, J. Leipzig, S. Hannenhalli, C. C. Berry, J. R. Ecker, and F. D. Bushman. 2006. Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog.* 2:e60.
- Llano, M., S. Delgado, M. Vanegas, and E. M. Poeschla. 2004. Lens epithelium-derived growth factor/p75 prevents proteasomal degradation of HIV-1 integrase. *J. Biol. Chem.* 279:55570–55577.
- Llano, M., D. T. Saenz, A. Meehan, P. Wongthida, M. Peretz, W. H. Walker, W. Teo, and E. M. Poeschla. 2006. An essential role for LEDGF/p75 in HIV integration. *Science* 314:461–464.
- Llano, M., M. Vanegas, O. Fregoso, D. Saenz, S. Chung, M. Peretz, and E. M. Poeschla. 2004. LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J. Virol.* 78:9524–9537.
- Malik, H. S., and T. H. Eickbush. 1999. Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. *J. Virol.* 73:5186–5190.
- Manns, A., M. Hisada, and L. La Grenade. 1999. Human T-lymphotropic virus type 1 infection. *Lancet* 353:1951–1958.
- Mitchell, R. S., B. F. Beitzel, A. R. Schroder, P. Shinn, H. Chen, C. C. Berry, J. R. Ecker, and F. D. Bushman. 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* 2:E234.
- Narezkina, A., K. D. Taganov, S. Litwin, R. Stoyanova, J. Hayashi, C. Seeger, A. M. Skalka, and R. A. Katz. 2004. Genome-wide analyses of avian sarcoma virus integration sites. *J. Virol.* 78:11656–11663.
- Nowrouzi, A., M. Dittrich, C. Klanke, M. Heinkelein, M. Rammling, T. Dandekar, C. von Kalle, and A. Rethwilm. 2006. Genome-wide mapping of foamy virus vector integrations into a human cell line. *J. Gen. Virol.* 87: 1339–1347.
- Oh, J., K. W. Chang, and S. H. Hughes. 2006. Mutations in the U5 sequences adjacent to the primer binding site do not affect tRNA cleavage by Rous sarcoma virus RNase H but do cause aberrant integrations in vivo. *J. Virol.* 80:451–459.
- Ozawa, T., T. Itoyama, N. Sadamori, Y. Yamada, T. Hata, M. Tomonaga, and M. Isobe. 2004. Rapid isolation of viral integration site reveals frequent integration of HTLV-1 into expressed loci. *J. Hum. Genet.* 49:154–165.
- Poeschla, E. M., F. Wong-Staal, and D. J. Looney. 1998. Efficient transduction of nondividing human cells by feline immunodeficiency virus lentiviral vectors. *Nat. Med.* 4:354–357.
- Pryciak, P. M., A. Sil, and H. E. Varmus. 1992. Retroviral integration into minichromosomes in vitro. *EMBO J.* 11:291–303.
- Pryciak, P. M., and H. E. Varmus. 1992. Nucleosomes, DNA-binding pro-

- teins, and DNA sequence modulate retroviral integration target site selection. *Cell* **69**:769–780.
48. **Puglia, J., T. Wang, C. Smith-Snyder, M. Cote, M. Scher, J. N. Pelletier, S. John, C. B. Jonsson, and M. J. Roth.** 2006. Revealing domain structure through linker-scanning analysis of the murine leukemia virus (MuLV) RNase H and MuLV and human immunodeficiency virus type 1 integrase proteins. *J. Virol.* **80**:9497–9510.
 49. **Schroder, A. R., P. Shinn, H. Chen, C. Berry, J. R. Ecker, and F. Bushman.** 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**:521–529.
 50. **Singleton, T. L., and H. L. Levin.** 2002. A long terminal repeat retrotransposon of fission yeast has strong preferences for specific sites of insertion. *Eukaryot. Cell* **1**:44–55.
 51. **Trobridge, G., R. K. Hirata, and D. W. Russell.** 2005. Gene targeting by adeno-associated virus vectors is cell-cycle dependent. *Hum. Gene Ther.* **16**:522–526.
 52. **Trobridge, G. D., D. G. Miller, M. A. Jacobs, J. M. Allen, H. P. Kiem, R. Kaul, and D. W. Russell.** 2006. Foamy virus vector integration sites in normal human cells. *Proc. Natl. Acad. Sci. USA* **103**:1498–1503.
 53. **Turlure, F., E. Devroe, P. A. Silver, and A. Engelman.** 2004. Human cell proteins and human immunodeficiency virus DNA integration. *Front. Biosci.* **9**:3187–3208.
 54. **Uchiyama, T.** 1997. Human T cell leukemia virus type I (HTLV-I) and human diseases. *Annu. Rev. Immunol.* **15**:15–37.
 55. **Withers-Ward, E. S., Y. Kitamura, J. P. Barnes, and J. M. Coffin.** 1994. Distribution of targets for avian retrovirus DNA integration in vivo. *Genes Dev.* **8**:1473–1487.
 56. **Wu, X., and S. M. Burgess.** 2004. Integration target site selection for retroviruses and transposable elements. *Cell Mol. Life Sci.* **61**:2588–2596.
 57. **Wu, X., Y. Li, B. Crise, and S. M. Burgess.** 2003. Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**:1749–1751.
 58. **Wu, X., Y. Li, B. Crise, S. M. Burgess, and D. J. Munroe.** 2005. Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J. Virol.* **79**:5211–5214.